Context-based image retrieval: A case study in background image access for Multimedia presentations

Sheng-Hao Hung, Pai-Hsun Chen, Jen-Shin Hong, Samuel Cruz-Lara

CONTEXT-BASED IMAGE RETRIEVAL: A CASE STUDY IN BACKGROUND IMAGE ACCESS FOR MULTIMEDIA PRESENTATIONS

[1]Sheng-Hao Hung, [2]Pai-Hsun Chen,[1,*]Jen-Shin Hong, and [3]Samuel Cruz-Lara
[1]*Department of Computer Science and Information Engineering, National ChiNan University, Taiwan*
[2]*Department of Computer Science and Information Engineering, Nan-Kai Institute of technology, Taiwan*
[3]*LORIA / INRIA Lorraine, Nancy-Université, France*
**Corresponding Author: E-mail jshong@ncnu.edu.tw*

ABSTRACT

Conventional approaches of image indexing and retrieval from digital libraries include content-based, metadata-based, and keyword-based approaches. This paper addresses a different way of image retrieval motivated by real-life applications for an intelligent system that can automatically select appropriate background images from textual passages. We explored techniques for developing automatic image-retrieval systems based on essential contextual information of a textual passage. We propose a framework that applies semantic role labeling techniques and a commonsense knowledge base, ConceptNet. The primitive results indicate that the proposed methodology has a potential on applications with textual passages that describe things and events that are regularly seen in every day life. However, for fantasy tales that describe truly fictitious things and events, the use of ConceptNet does not allow to obtain accurate results.

KEYWORDS

Context-based image retrieval, knowledgebase, semantic role labeling, ConceptNet, Multimedia.

1. INTRODUCTION

Image indexing and retrieval from digital libraries have been extensively studied for decades. In the literature, there has been a broad variety of approaches proposed to retrieve images efficiently for end users. A content-based image retrieval (CBIR) approach (Smeulders et al. 2000) relies on certain low-level image features, such as color, shape, and texture, for retrieving images. A major drawback of this approach is that there is a "semantic gap" between low-level features of images and high-level human concepts (Smeulders et al 2000). Therefore, at this stage, the applicability of the content-based image retrieval methods is still limited. For metadata-based image retrieval, an image item is annotated with descriptions which form a database "record" to represent it within an image archive. A record is generally further divided into various metadata fields (elements), each one representing a particular type of information that help to identify it. The efficiency of the indexing and precision of such retrieval is in general acceptable with current database technologies. However, the major problem in this case is the difficulty of designing an "optimal" structure of the term-based fields. Keeping the efficiency of retrieval, while keeping the annotation efforts affordable to cover a wide spectrum of semantics in an image collection, is not an effortless task. So, many existing archives further incorporate the "keyword" field in the metadata model, therefore end users can search images using a keyword-based search.

This paper addresses a different way of image retrieval. We intend to develop automatic systems that search images from a textual passage. An example of real-life applications is to allow a Blog-system to automatically select appropriate background images for textual passages written by users. For example, given a passage describing "bikini", "volley ball", and "surfing", we would expect the system to choose an appropriate background image, such as a beach. In the following sections, we will describe our methodology. This methodology is based on a commonsense knowledge-base, ConceptNet, and a natural language processing technologies, the semantic role labeling technique.

2. METHODOLOGY

In this section, first, we will briefly describe the essential technologies applied. Then, we will elaborate the underlying principles and the operations of each associated stage.

2.1 Semantic Role Labeling (SRL) Techniques

Roughly speaking, in a sentence, a verb (predicate) indicates an event. A semantic role is the relationship that a syntactic argument has with the verb. One of the most commonly-used schemes for specifying the semantic roles is to construct a large-scale corpus such as the PropBank (Palmer et al. 2005). Within PropBank, the arguments of a verb are labeled sequentially from ARG0 to ARG5, where ARG0 is usually the subject of a transitive verb; ARG1, its direct object, and so on. A variety of adjunctive arguments, such as ARGM-LOC for locatives, and ARGM-TMP for time, are also tagged. As an illustrative example, the semantic roles for the sentence "I saw a girl in the park in the morning" based on the PropBank style markup are:

[ARG0 **I**] [Target **saw**] [ARG1 **a girl**] [ARGM-Loc **in the park**] [ARGM-Tmp **in the morning**]

Semantic Role labeling techniques automatically identify the semantic roles of a sentence. In the literature, there are a number of studies proposing different methodologies for such purposes, for example: (Gildea and Jurafsky 2002; Pradhan et al. 2004; Koomen et al. 2005). These methodologies have obtained well accurate results about 80% on ARG0, ARG1, and 70% on ARGM-LOC, ARGM-TMP, for some sample data coming from the Wall Street Journal (Carreras and Márquez 2005; Pradhan et al. 2004). In (Lin et al. 2007), we have applied these SRL techniques to detect event-based knowledge in digital image descriptions with satisfactory precision.

2.2 ConceptNet: A commonsense knowledgebase

ConceptNet is a knowledgebase developed in the OMCS project (Singh et al. 2002). It is a freely available semantic network consisting of more than 200,000 elements of commonsense knowledge contributed by thousands of volunteers across the Web. ConceptNet includes a wide range of commonsense concepts interlinked by about twenty semantic relations, such as "EffectOf", "IsA", and "UsedFor". The knowledge structure allows a variety of primitive inferences such as temporal, spatial, and affective. The abundance of the semantic relations supported in ConceptNet makes it of practical value in many applications on intelligent systems, for example: (Liu and Singh 2002; Liu et al. 2003).

2.3 Frameworks

The overall framework for our context-based background image retrieval system includes three major stages:

a) **Recognize the semantic roles.**
In the first stage, the sentences within text passages are processed using SRL tools in order to identify the semantic roles. At this stage, we have chosen what we considered thee three most essential semantic roles for the background image retrieval applications, including "subject", "object" and "location". These semantic roles are then used within the next stage.

b) **Retrieve Contextual knowledge from ConceptNet.**
The relations in ConceptNet includes "IsA", "propertyOf", "partOf", "MadeOf", and "OftenNear", "LocationOf", "UsedFor", "CanDo", "CapableOf" , "FirstSubeventOf", "LastSubeventOf", "EffectOf", "EventForGoalEvent", "DesiresEvent", "EffecOfIsState", "EventRequiresObject", "DesiresNotEvent". At this stage, we have chosen what we considered the two most relevant relations, namely, the "OftenNear" and "LocationOf" relations, as the main components of the contextual knowledge for the background image retrieval applications.

c) **Use context-base keywords to query Image databases.**

Next, a "query formulation module" is required to transform information of the core semantic roles and contextual knowledge of a passage into effective query strings in order to retrieve images from keyword-based digital image archives. The query formulation module adopts an adaptive and iterative strategy that progressively refines the query strings so that the retrieved images could best fit into the context of the passages. For example, consider the following passage in a child story book:

"I see three lemurs jumping around and screaming. They are scared by the snake. However, a sloth is still soundly sleeping. Around the corner, many children are watching a shark swimming swiftly."

The subjects and objects in this passage include "snake", "lemur" and "sloth". No information is provided for ARG-Loc. During the first execution of the image retrieval process, the query string is formulated as "snake AND lemur AND sloth". In the case the response from the image archives indicates that there is no image annotated with all these terms, there is a need for a second query. Presumably, one reasonable strategy is to find images of a place where the "snake", "lemur", and "sloth" could all possibly appear. Such information can be obtained by using the intersection of the "locationOf" in ConceptNet for "snake", "lemur", and "sloth". A query to the ConceptNet for this intersection gives a number of most frequently places such as "jungle", "forest", and "zoo". In principle, since there is no other information inside the related textual passages, images of "jungle", "forest", and "zoo" can be considered as reasonable background images for the passage. If another sentence, such as "A shark is swimming rapidly," appears within the textual passage, a more reasonable background image will be obtained by using the intersection of the "locationOf" in ConceptNet for "snake", "lemur", "sloth", and "shark". In such case, the only location obtained from ConceptNet is "zoo" which should be used as the query string to access the images.

In general, depending on the genre and relative degree of importance of the semantic roles and contextual knowledge obtained for a textual passage, different heuristic strategies have to be designed to find the optimal query mechanisms to access the desired images. For example, in a situation where a clearly shown "ARG-LOC" appeared in the passage, very likely the events described in the passages occurs at the given location of the ARG-LOC. In that case, the ARG-LOC could be possibly used as the keywords to query the images. To develop a more robust system, based on the empirical studies using real-life data and image archives, we have manually designed a variety of heuristic strategies to handle different situations.

3. TEST RUNS AND DISCUSSION

In order to verify the applicability of the proposed methodology within real-life applications, we have conducted a primitive test run experiments. We use an automatic semantic labeling engine - ASSERT (http://oak.colorado.edu/assert/) to parse the semantic roles of sentences in textual passages. A prototype image retrieval system implemented using ASP.Net based on the methodology proposed has also been developed. This system has been used in order to perform test runs. As an example of the test runs, we consider three example passages that have been accumulated from some web pages that we have used to query a private collection of images annotated with keywords.

Table 1 shows the passages, the semantic roles and the keywords that have been obtained. For the first three passages, the obtained keywords appear to have a good match with respect to the events described in the passages. For passages having a more complicated syntax, such as the fourth passage shown in Table 1, ASSERT failed to give an accurate parsing on the semantic roles. Furthermore, in this case, based on the subjects and objects parsed, the ConceptNet failed to provide any location where these subjects and objects could all be located.

In this stage, we are currently working on the further refinement on the heuristic strategies for mapping the semantic information gathered in the passage to the query strings for the keyword-based image archives.

The primitive results we have obtained indicate that the proposed methodology has a potential on the application of background image retrieval for passages that describe things and events that are regularly seen in every day life. However, for passages having more complicated syntax, such as fantasy tales that describe truly fictitious things and events, it seems that the direct applicability of ConceptNet does not produce accurate results. After all, as compared to the scope of general human knowledge, the number of data entries in ConceptNet is far limited. Currently, we are also working on techniques that apply web-based text mining techniques (Etzioni et al. 2005; Girju et al. 2006) to effectively and efficiently increase the coverage of the knowledgebase.

Passages	Parsed Semantic Roles				Keywords
	Subject	Object	Location	Time	
Seals are endangered in Hawaii. Their pups died without parents care. They are often eaten by sharks. A team wants to save the seals. They propose to kill 10 sharks this summer.	seals pups team	seals sharks	Hawaii	summer	ocean sea
The camel is its own miracle. It could bite cactus to deal with lack of water. It is as the caravan oasis. It is just like grapes grown in the sand.	camel cactus oasis grapes				desert
Summer is here. Girls always wear bikini. The young like to play beachball. But be careful! You need sunscreen against the sunburn.	bikini beachball sunscreen sunburn				beach
The Queen was jealous, and ordered a huntsman to take Snow White into the woods to be killed. She demanded that the huntsman return with Snow White's lungs and liver as proof. The huntsman took Snow White into the forest, but found himself unable to kill the girl. Instead, he let her go, and brought the queen the lungs and liver of a wild boar.	Queen huntsman	Snow White lung liver girl Queen			

Table 1. Passages, parsed semantic roles, and keywords in the example test runs of the system.

4. CONCLUSIONS AND FUTURE WORK

This study reports a primitive work on a framework for context-based image retrieval that applies semantic role labeling techniques and a commonsense knowledgebase, ConceptNet. Our approach is motivated by real-life applications for an intelligent system that can automatically select appropriate background images for textual passages written by users. As we have shown in the test runs, the proposed methodology has a potential on the application of background image retrieval for passages that describe things and events that are regularly seen in daily lives. However, when using fictional textual information, as in fantasy tales for example, the direct applicability of ConceptNet is restricted. One possible reason for this is that the events described in many fictional passages are often against human commonsense. Of course, one may take into account that ConceptNet is very limited with respect to the scope of general human knowledge, so if the number of entries in the applied knowledgebase is increased, one may expect to obtain more accurate results. Another possible interesting way to effectively and efficiently increase the coverage of the knowledgebase, is to apply web-based mining text techniques, this is the work we are currently doing.

Finally, at LORIA / INRIA Lorraine, the "Multi Lingual Information Framework" MLIF [ISO AWI 24616] is being designed with the objective of providing a common conceptual model and a platform allowing interoperability among several translation and localization standards. So, in the framework of future work activities and with the assistance of MLIF, we intend to apply the proposed approach to other languages, such as French, Chinese, Spanish.

ACKNOWLEDGEMENT

This work is supported by the National Science Council of Taiwan, under the grant numbered as NSC-94-2422-H-260-002.

REFERENCES

Carreras and Márquez 2005. Introduction to the CoNLL-2005 shared task: Semantic role labeling. *In proceedings of the CoNLL-2005*, pp. 152–164.

Etzioni et al. 2005. Unsupervised named-entity extraction from the Web: an experimental study, *Artificial Intelligence*, Vol. 165, pp.91-134.

Gildea and Jurafsky 2002. Automatic labeling of semantic roles. *Computational Linguistics*, Vol. 28 No.3, pp. 245-288

Girju et al. 2006. Automatic discovery of part-whole relations, *Computational Linguistics*, Vol.32, No.1, pp.83-135.

Koomen et al. 2005. Generalized inference with multiple semantic role labeling systems. *In Proceedings of the Ninth Conference on Computational Natural Language Learning,* CoNLL-2005, pp. 181-184.

Lin et al. 2007. Semantic Role Labeling Techniques for Event-based Knowledge Extraction from Free-text Descriptions for Art Images. *The Electronic Library*, v26, n1, scheduled to appear in April 2008.

Liu and Singh 2002. MAKEBELIEVE: Using commonsense to generate stories. *Proceedings of the 20th National Conference on Artificial Intelligence, (AAAI-02)*, Student Abstracts, pp. 957-958, Seattle, WA.

Liu et al. 2003. A model of textual affect sensing using real-world knowledge. *Proceedings of the Seventh International Conference on Intelligent User Interfaces (IUI 2003)*, pp. 125-132. Miami, Florida.

Palmer et al. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, Vol. 31 No. 1, pp. 71-106.

Pradhan et al. 2003. Semantic role parsing: adding semantic structure to unstructured text. In *Proceedings of the International Conference on Data Mining (ICDM 2003)*.

Singh et al. 2002. Open mind common sense: knowledge acquisition from the general public, *Proceedings of the First International Conference on Ontologies, Database, and Applications of Semantics for Large Scale Information Systems*. Lecture Notes in Computer Science, Vol. 2519, Heidelberg: Springer-Verlag.

Smeulders et al. 2000. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 22 No. 12, pp. 1349-1380.